Mathematical, Information and Computational Sciences

Computer Science PI Meeting
June 26-27, 2002

Fred Johnson

WELCOME!!

- To The First (annual?) MICS Computer Science PI Meeting
  - Foster a sense of community in CS focused on high end problems
  - Get to know each other and what's going on
  - Future directions roadmap
- Program Managers
- Folks who made it happen!
  - Cheryl, Bonnie, Jim, Rusty

Mission

Mathematical, Information and Computational Sciences

Discover, develop, and deploy the computational and networking tools that enable researchers in the scientific disciplines to analyze, model, simulate, and predict complex physical, chemical, and biological phenomena important to the Department of Energy (DOE).

foster and support fundamental research in advanced scientific computing – applied mathematics, computer science, and networking

operate supercomputers, a high performance network, and related facilities.

FY 2003 President's Request
Advanced Scientific Computing Research Program

Mathematical, Information and Computational Sciences

Budget Authority
($ in thousands)

	FY2001	FY2002	FY2003
Mathematical, Information, and Computational Sciences	$151,647	$154,400	$166,625
Laboratory Technology Research	$ 9,649	$ 3,000	$ 3,000
TOTAL ASCR	$161,296	$157,400	$169,625

NOTE- FY2001 excludes SBIR/STTR set-asides

Program Strategy

Mathematical, Information and Computational Sciences

Fundamental Research → R&D for Applications → Testbeds → *BES, BER, FES, HEP, NP*

Research to enable...

...simulation of complex systems

- Applied Mathematics
- Computer Science
- Advanced Computing Software Tools
- Scientific Application Pilots
- Integrated Software Infrastructure Centers

SciDAC!

Teams- mathematicians, computer scientists, application scientists, and software engineers

SciDAC!

- Applications
  - Materials
  - Chemical
  - Combustion
  - Accelerator
  - HEP
  - Nuclear
  - Fusion
  - Global Climate
  - Astrophysics

...distributed teams and remote access to facilities

- Networking
- Collaboratory Tools
- Collaboratory Pilots

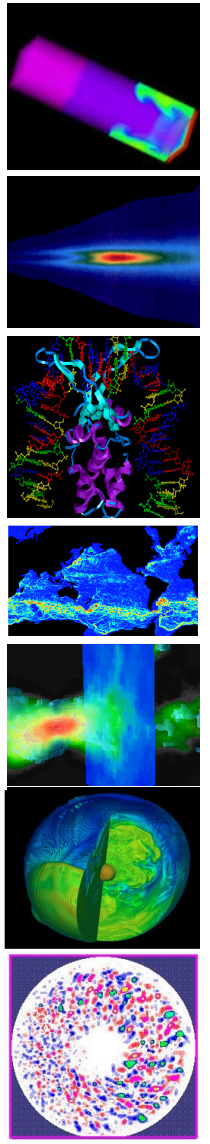
SciDAC!

- Facility Access
- Teams of Researchers

High Performance Computing and Network Facilities for Science

- National Energy Research Scientific Computing Center (NERSC)
- Advanced Computing Research Facilities
- Energy Sciences Network (ESnet)

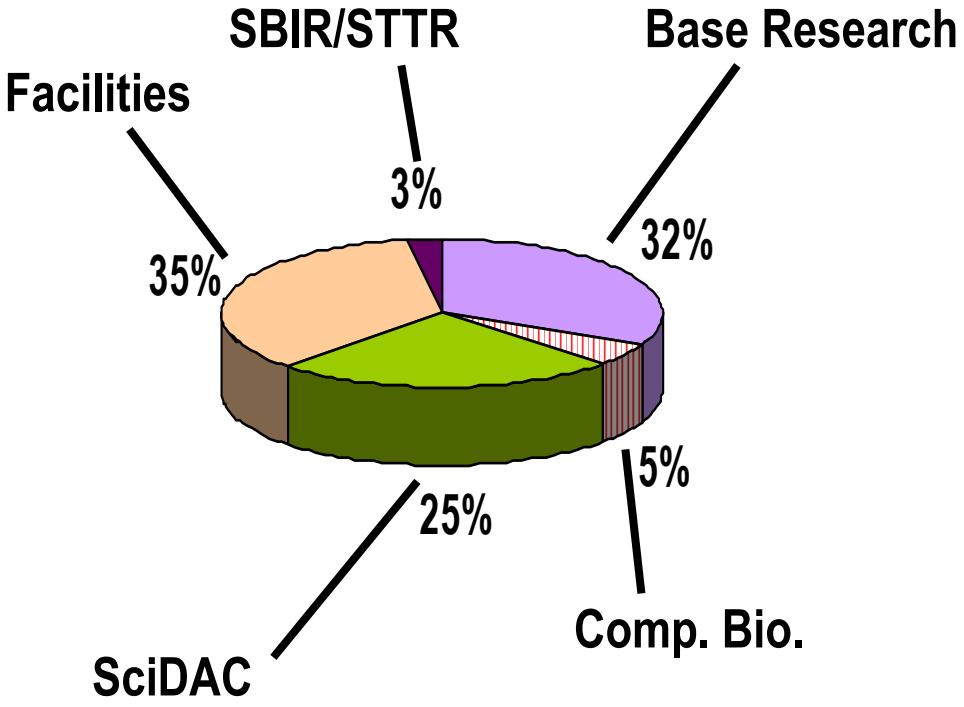
SciDAC!



Budget Request

Mathematical, Information and Computational Sciences

FY2003- $166,625,000



Enhancements over FY2002

- **Computational Biology +$5.6M**
- **SciDAC +$5.3M**
- **Facilities +$1.3M**

FY2001 MICS Research Budget by Institution
$ in thousands (# of projects)

Mathematical, Information and Computational Sciences

	Base Research						
	AMS	CS	NC-ACST	NR	SAPP	SciDAC	Comp. Bio.
Univ. (& Others)	8,236	9,336	5,597	2,583	1,105	17,548	1,703
	(42)	(24)	(20)	(12)	(10)	(56)	(8)
Laboratories	15,496	11,605	12,984	2,673	960	19,895	1120
	(31)	(44)	(74)	(16)	(7)	(65)	(3)
Totals	23,782	20,941	18,581	5,256	2,065	37,443	2,823

Legend

AMS- Applied Mathematical Sciences
CS- Computer Sciences
NC-ACST- National Collaboratories- Advanced Computing Software Tools
NR- Networking Research
SAPP- Scientific Application Pilot Projects
SciDAC- Scientific Discovery through Advanced Computing
Comp. Bio.- Computational Biology

Program Evolution

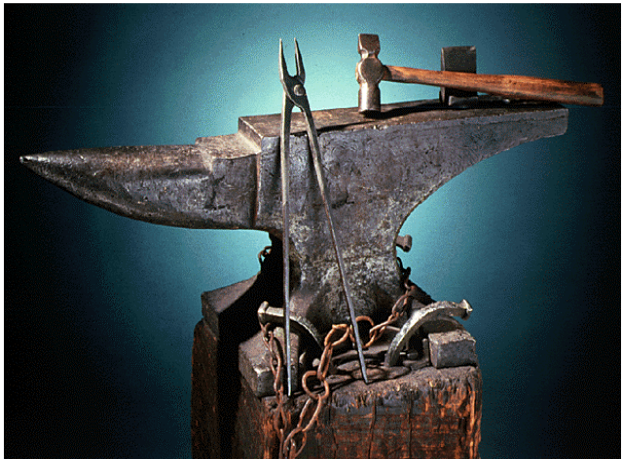
Mathematical, Information and Computational Sciences

- **FY 2001**
  - Initiated software infrastructure component of SciDAC
  - Initiated research efforts in computational biology
  - Upgraded NERSC to 5 teraflops
  - Acquired IBM Power 4 Hardware for evaluation/scaling (limited SciDAC support)
- **FY 2002 Activities**
  - Ensure success of SciDAC
  - Strengthen base research effort (Early Career PI)
  - Initiate NERSC-4 acquisition process
- **FY 2003 Plans**
  - Launch computational component of Genomes to Life, in partnership with BER
  - Initiate computational nanoscience partnership with BES as part of SciDAC
  - Provide topical high performance computing resources to support SciDAC research

Computer Science Research

Mathematical, Information and Computational Sciences

- **Challenge – HPC for Science is (*still after fifteen years!*)**
  - Hard to use
  - Inefficient
  - Fragile
  - An unimportant vendor market
- **Vision**
  - A comprehensive, integrated software environment that enables the effective application of high performance systems to critical DOE problems
- **Goal– Radical Improvement in**
  - Application Performance
  - Ease of Use
  - Time to Solution



System Admin	Software Development	Scientific Applications
Res. Mgt	Framewrks	PSEs
Scheduler	Compilers	Viz/Data
Chkpt/Rstrt	Debuggers	Math Libs
File Sys	Perf Tools	Runtme Tls
User Space Runtime Support		
OS Kernel		OS Bypass
Node and System Hardware Arch		

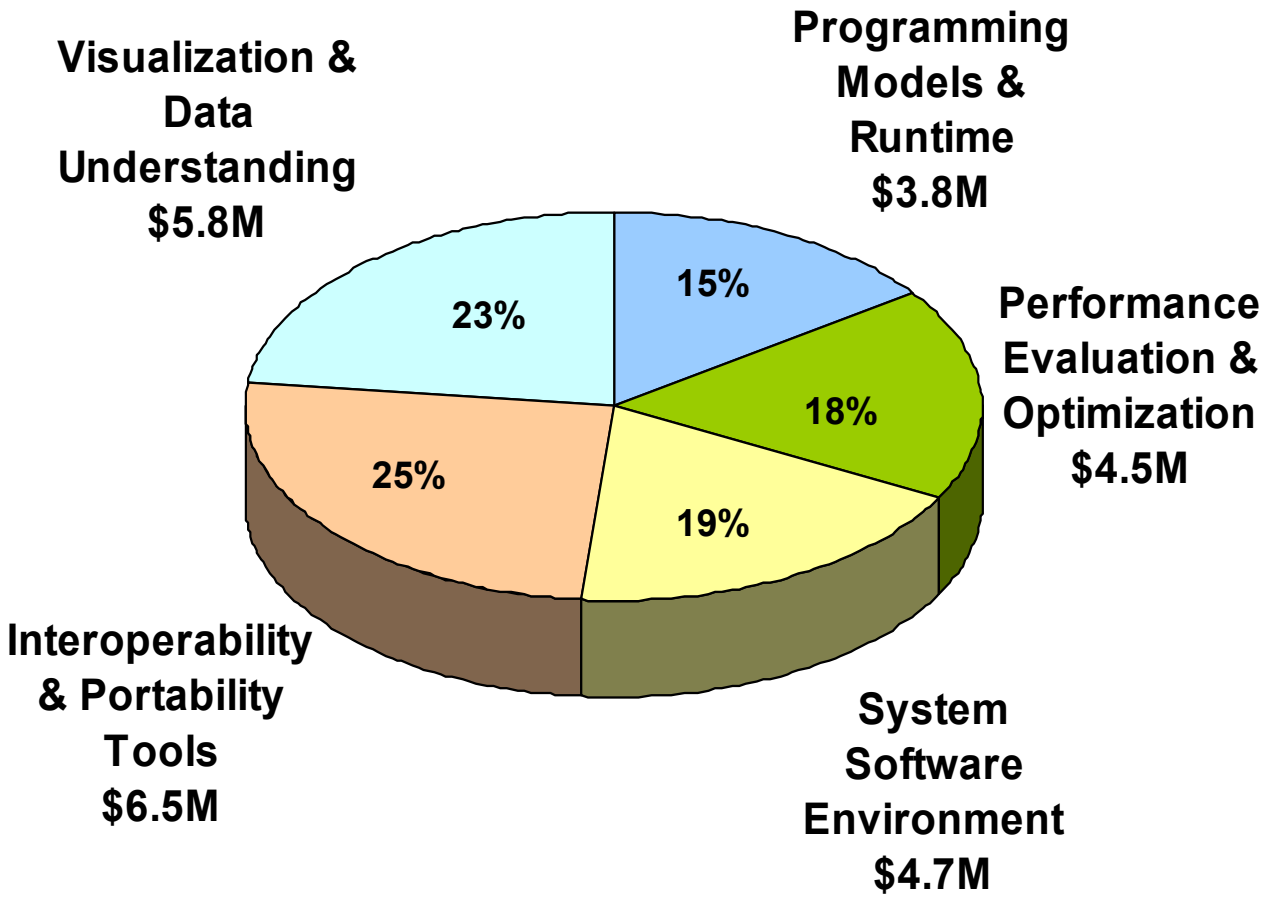
HPC System Elements

Program Components

- **Base Program**
  - Evolutionary and revolutionary software methodologies for future generations of HPC architectures
- **SciDAC Integrated Software Infrastructure Centers**
  - Enable effective application of current terascale architectures to SciDAC applications through focused research and partnerships

Computer Science Technical Elements

Mathematical, Information and Computational Sciences



Opportunities for Program Growth

Mathematical, Information and Computational Sciences

- Dynamic OS/Runtime environments
- Operating systems for petascale systems
- Application specific problem solving environments
- Intelligent program development environments
- Accelerate HW/SW for effective petascale computation
  - **Life Science (such as those described in the Genome-to-Life initiative)**
  - **Nanoscience (such as those proposed in the NSF and DOE Nanoscience initiatives)**
  - **Computational Cosmology and Astrophysics**
  - **Earth Science and Environmental modeling**
  - **Computational Physics**
  - **Computational Chemistry**
  - **Fusion modeling and simulation**
  - **Multidisciplinary design problems**

Scientific Discovery through Advanced Computing (SciDAC)

Mathematical, Information and Computational Sciences

An ***integrated*** program to:

(1) Create a new generation of **Scientific Simulation Codes** that take full advantage of the extraordinary computing capabilities of terascale computers.

(2) Create the **Mathematical and Computing Systems Software** to enable the Scientific Simulation Codes to effectively and efficiently use terascale computers.

(3) Create a **Collaboratory Software Environment** to enable geographically separated scientists to effectively work together as a team and to facilitate remote access to both facilities and data.

SciDAC Program Elements

Mathematical, Information and Computational Sciences

- **Scientific Simulation Codes**
  - Funding in ASCR, BES, BER, FES, and HENP to develop scientific codes that take full advantage of terascale computers
- **Mathematical Methods and Algorithms**
  - Funding in ASCR to develop mathematical methods and algorithms that perform well on cache-based microprocessors and scale to thousands and, eventually, tens of thousands of processors
- **Computing Systems Software**
  - Funding in ASCR to develop software to facilitate the development and use of scientific codes for terascale computers

Program Elements
(cont'd)

- Funding in ASCR to develop software needed to manage and analyze massive data sets produced by simulations on terascale computers

- **Collaboratories and Data Grids**
  - Funding in ASCR to develop collaborative software to enable geographically-dispersed researchers to work as a team
  - Funding in ASCR to develop computational and data grids to facilitate access to computers, facilities, and data

Scientific Simulation Methods and Codes for Terascale Computers

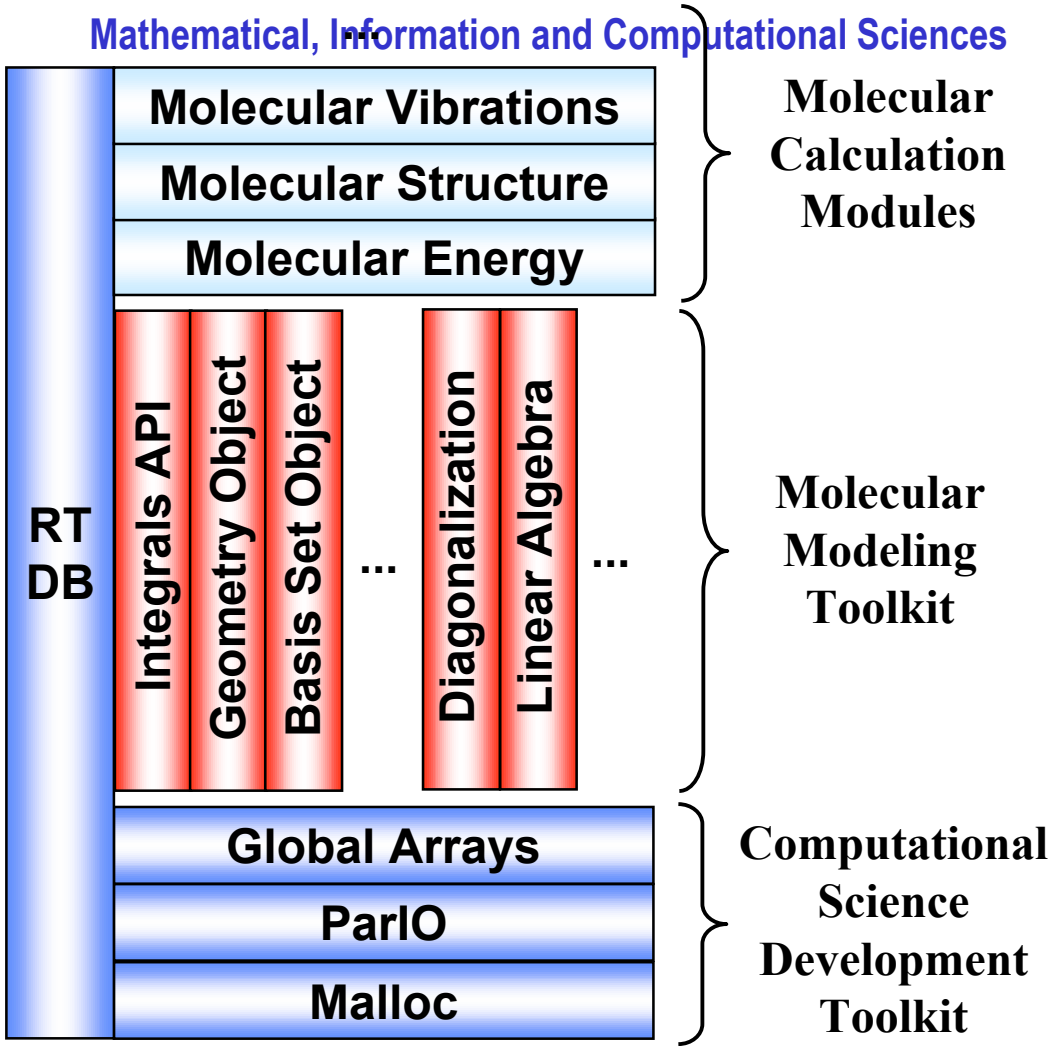
Mathematical, Information and Computational Sciences

- **BES ($1,931)**
  - Understand and predict the energetics and dynamics of chemical reactions and the interaction between chemistry and fluid dynamics – electronic structure and reacting flows
- **BER ($8,000)**
  - Understand and predict the earth's climate at both regional and global scales for decades to centuries, including levels of certainty and uncertainty
- **FES ($3,000)**
  - Understand and predict microscopic turbulence and macroscopic stability in magnetically confined plasmas, including their effect on core and edge confinement

Scientific Simulation Methods and Codes for Terascale Computers (*cont'd*)

  - Understand and predict the electromagnetic fields, beam dynamics, and other physical processes in heavy-ion accelerators for inertial fusion
  - Understand basic plasma science processes, such as electromagnetic wave-particle interactions and magnetic reconnection
- HENP ($7,000)
  - Understand and predict electromagnetic field and beam dynamics in particle accelerators
  - Understand and predict the physical phenomena encompassed in the Standard Model of Particle Physics
  - Understand mechanisms of core collapse supernovae

SciDAC Prototype: NWChem



NWChem

a major new modeling capability for molecular science

Molecular Electronic Structure
Molecular Dynamics
600,000 lines of code and growing

Runs on ...

Cray T3D/E, IBM SP2, SGI SMP, NOWs, Sun and other workstations, X86 PCs (Linux)

Scales to ...

2,000+ processors

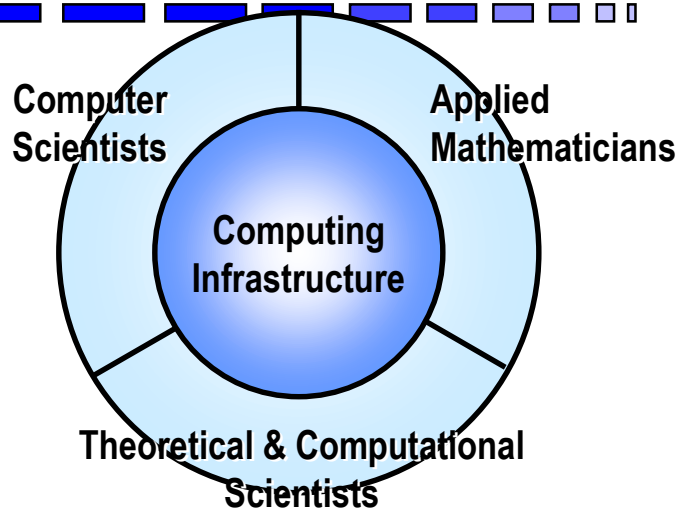
Developers ...

Core group (15) plus larger group (20) of world-wide collaborators
100 person-years at PNNL alone

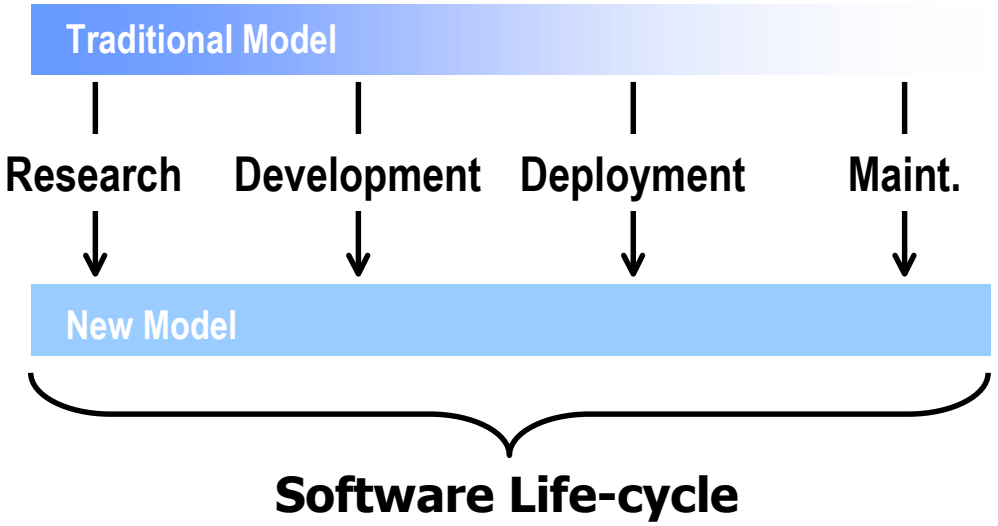
Emphasis on TeamBuilding and Complete Software Life-Cycle

Mathematical, Information and Computational Sciences

1 **Bring together teams** of theoretical and computational scientists, computer scientists, applied mathematicians, with the computing infrastructure, and **establish close working relationships**.



2 **Support and integrate research** through **development** to **deployment** to ensure that the scientific community receives innovative, yet usable software capabilities in a timely fashion; seek industrial support wherever possible.



SciDAC Program Awards -- Integrated Software Infrastructure Centers (ISIC)

Mathematical, Information and Computational Sciences

ISIC Vision:

- Provide a comprehensive, portable, and fully integrated suite of systems software, libraries and tools for the effective management and utilization of terascale computers by SciDAC applications
- Provide maximum performance, robustness, portability and ease of use to application developers, end users and system administrators

Award Summary:

- 3 Centers -- Mathematical algorithms/libraries, $8.6M/year
- 4 Centers – Computer science issues, $10.7M/year

Computer Science Integrated Software Infrastructure Centers

Mathematical, Information and Computational Sciences

Four activities focused on a comprehensive, portable, and fully integrated suite of systems software and tools for effective utilization of terascale computers - $10.7M

Scalable Tools for Large Clusters; Resource Management; System Interfaces; System Management Tool Framework	ANL LBNL Ames PNNL SNL LANL National Center for Supercomputing App (Al Giest, ORNL)	$2.2M/Yr
High-End Computer Sysetms Performance: Science & Engineering	ANL LLNL ORNL U. of Illinois UCSD U. of Tennessee U. of Maryland (David Bailey, LBNL)	$2.4M
High Performance, Low Latency Parallel Software Component Architecture	ANL LANL LLNL ORNL PNNL U. of Utah Indiana U. (Rob Armstrong, SNL)	$3.1M
Scientific Data Management Enabling Technology	ANL LLNL ORNL Georgia Tech. UCSD Northwestern U. North Carolina State (Ari Shoshani, LBNL)	$3M

Open Source/IP

- Strong emphasis on Open Source
- One license doesn't fit all
- Code from many sources needed
- Independent IP holding organization
  - CCA subset – SNL, ANL, Indiana
  - Goal: Code development without FEAR

The Future of Computational Science at DOE

Mathematical, Information and Computational Sciences

From Secretary Abraham's speech at BNL on Friday, June 17, 2002

The Department is also one of this nation's major sponsors of advanced computers for science. We did this in the first instance for obvious national security reasons. And we have gone on to establish the country's first supercomputer center for science. Now more than ever, however, virtually all science depends on teraflops. The computer is no longer simply a tool for science. Computation is science itself, and enables scientists to understand complex systems that would otherwise remain beyond our grasp. It's an indispensable contributor to our national security work, to nanotechnology, as well as to every other venture we undertake in science. **And I intend that this Department maintains America's lead in this critical field.**

Today's Agenda

- **Morning**
  - Bob Lucas – Lessons Learned the Hard Way
  - Tony Mezzacappa – Software Perspectives of a SciDAC Application Scientist
  - Burton Smith – New Ideas, New Architectures
- **Afternoon**
  - ISIC Posters
  - Base Program Posters

Thursday's Agenda

- Morning Breakout Sessions
  - Programming Models and Runtime
  - System Software Environment
  - Interoperability and Portability Tools
  - Visualization and Data Understanding
  - If you don't like these …
- Afternoon
  - Feedback by session chairs
  - General discussions, meeting feedback, …